Replication-Dependent Mechanism of Chromosome Fragility at the Site of Inverted Repeats in *Saccharomyces cerevisiae*

Clara Moon
School of Biology
Georgia Institute of Technology

Advisor: Dr. Kirill Lobachev
Reader: Dr. Yury Chernoff

December 5, 2006
Honors Thesis

ABSTRACT

Genome instability is linked to cancer and many hereditary diseases. Chromosomal aberrations are often associated with repeats that can adopt DNA secondary structures. Studying the mechanism of genome instability caused by unstable motifs will therefore contribute to our understanding of the origin of human pathology. In this study, we investigate the mechanism underlying inverted repeated associated instability. A biological assay was used to identify mutant strains that showed increased instability in the presence of inverted repeats in the Δ*sae2* or Δ*mre11* background. Tetrad analysis techniques were used to map the mutations, and four mutations (all isolated independently) were identified in the *CDC2* (*POL3*) gene. We focused on one mutation, *pol3-6PL* for this study. The effect of *pol3-6PL* on inverted repeat-mediated instability was assessed using armloss and gene amplification assays. Fluctuation tests on the mutant strain showed increased rates of these events. CHEF and Southern blot hybridization showed increased double-strand break formation at the site of inverted repeats. From this data, we propose a model for a replication-dependent mechanism of chromosomal fragility at the site of inverted repeats.

INTRODUCTION

Repetitive Sequences and Secondary Structures

Eukaryotic genomes are enriched with repetitive sequences in the DNA, including dinucleotide repeats, trinucleotide repeats, purine-pyrimidine mirror repeats, and palindromic sequences (or inverted repeats). These repetitive sequences can form a variety of noncanonical secondary structures such as hairpins, cruciforms, and H-DNA.

These secondary structures can alter the DNA architecture as well as interfere with normal DNA processes, such as replication and transcription events (Bissler, 1998). These structures in the genome have been associated with human diseases (as discussed below).

One area of particular interest is the inverted repeat, or palindromic sequence. A palindrome can be defined as a sequence that is identical in the 5' to 3' direction in both strands of the DNA. Under favorable conditions, Watson-Crick base pairs can form between the bases of the inverted repeat within one strand of the DNA to form hairpin and cruciform structures. The presence of single-stranded DNA promotes the formation of secondary structures in the form of hairpins (Lewis and Cote, 2006). Local changes in superhelicity leading to unwinding torsional stress (or negative supercoiling) can relax the DNA double helix, open up a small single-stranded region in the DNA, and promote base-pairing within each strand to form cruciform structures (Sinden, 1994).

Inverted repeats sequences do not have to be perfect palindromes to form secondary structures. A spacer region of different sizes can exist between the two inverted repeats, and the palindrome repeats themselves may have base-pair differences. However, increasing the spacer length and/or increasing sequence divergence both lead to a decreased probability of forming a stable secondary structure (Lobachev et al., 2000). In addition, the length of the inverted repeat itself contributes to the stability of the secondary structure formed.

Double-Strand Breaks, Homologous Recombination, and Genomic Instability

Research performed in both prokaryotic and eukaryotic systems have demonstrated instability associated with palindromes. Palindromes in bacterial cells, yeast cells, and mammalian cells have all been shown to lead to rearrangements and deletions (Leach, 1994; Lewis et al., 1999; Lobachev et al., 2002; Narayanan et al., 2006). Palindrome-associated genomic instability has been attributed to the ability of these sequences to adopt secondary structures.

Long palindromic sequences in bacteria were shown to be deleted at very high rates, and propagation of these repeats was difficult to carry out in wild-type strains (Leach, 1994 and references therein). The identification of *sbcC* and *sbcD* mutants in *E. coli* allowed for the propagation of long palindromes in bacterial systems (Leach and Stahl, 1983; Lloyd and Buckman, 1985). Further studies on SbcC and SbcD helped to elucidate their role in the cell. It was found that SbcCD has both double-strand exonuclease and single-strand endonuclease activity (Connelly and Leach, 1996). Connelly and Leach proposed that SbcCD was involved in cleaving hairpins that formed during replication. Once hairpins are cleaved, homologous recombination would be carried out by the RecABCD complex, preventing the collapse of the replication fork as well as mutagenesis from attempted bypass of the hairpin. In the presence of long palindromic repeats, DNA replication would be inhibited due to the continued cycle of cleavage and recombination by these two complexes. However, in *sbcCD recBC* strains, replication would be allowed to continue, allowing for the propagation of these unstable inverted repeats.

Though prokaryotic cells have long been used to study instability created by palindromes, these repetitive sequences have been studied in eukaryotic cells as well. One such model organism that is commonly used is *Saccharomyces cerevisiae*, or the budding yeast. *S. cerevisiae* is one of the simplest eukaryotes. This single-cellular organism is easily transformed with foreign DNA, can exist in both a diploid and haploid state (which simplifies certain genetic manipulations), and grows rapidly (Sherman, 2002). The genome consists of 16 chromosomes and has been completely sequenced. They are non-pathogenic and require less maintenance than most of the other higher eukaryotic organisms. These and other qualities make yeast an ideal model organism to perform genetic studies.

The *S. cerevisiae* homologs of SbcD and SbcC are Mre11 and Rad50 respectively (Connelly et al., 1999). In yeast, these two proteins are found to be a part of the MRX complex (comprised of Mre11, Rad50, and Xrs2). Just like their prokaryotic homologs, Mre11 and Rad50 were shown to have endonuclease and exonuclease activities. It was shown *in vitro* that the MRX complex is involved in cleaving and processing hairpin structures formed during replication (Trujillo and Sung, 2001). All evidence seemed to point towards this hypothesis, however, physical detection of DSBs was not shown in these works, allowing the possibility for the MRX complex to be involved in another pathway.

In further studies with *S. cerevisiae*, it was shown that inverted *Alu* repeats lead to unique hairpin-capped DSBs (Lobachev et al., 2002). *Alu* repeats are repetitive sequences that are commonly found in clusters in the human genome (Lobachev et al., 2000). These sequences are often found in the inverted repeat orientation. Therefore, studying the

effects of instability caused by these repeats could have implications for human genetic diseases.

Spontaneous DSBs that occur in the genome have exposed 5' and 3' termini (Fig. 1A). DSBs occurring at the site of inverted repeats, however, were found to have hairpin-capped ends (Fig. 1B) (Lobachev et al., 2002). The proposed model for the formation of these breaks is that the palindromic sequence leads to the formation of a cruciform structure, which is cleaved by an enzyme, leading to a DSB in the DNA. A ligase (that may or may not be a part of the resolvase that cleaves the cruciform) ligates the breaks that are formed to create hairpin-capped ends. The proposed model is shown below; however some details (such as the specific enzyme(s) involved in cruciform resolution) are still unclear.

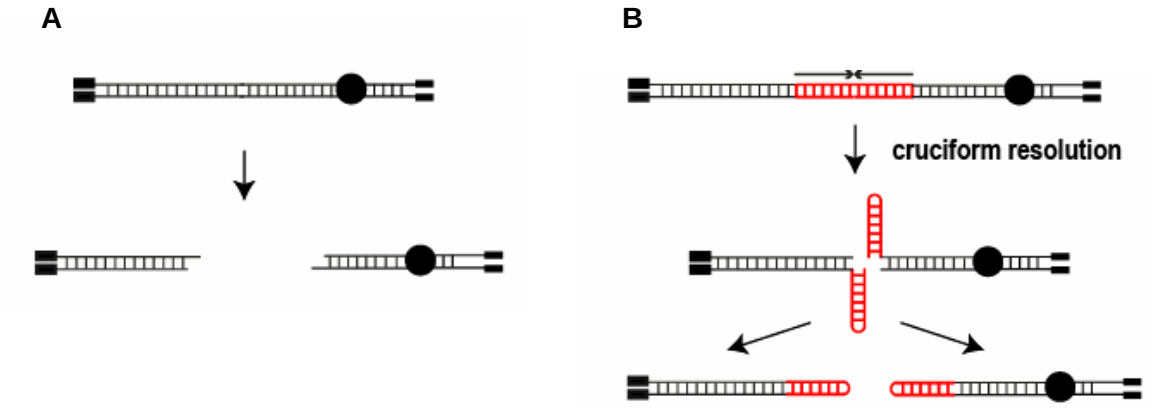

Fig. 1. Different consequences of double-strand breaks. (A) Spontaneous formation of DSBs in the genome. These breaks lead to exposed 5' and 3' termini. (B) Model of the formation of hairpin-capped broken molecules due to DSBs that occur in the presence of inverted repeats. Figure adapted from (Lobachev et al., 2002).

In addition, Lobachev et al. showed that the MRX complex is not involved in cleaving hairpin structures during replication, but is instead involved in processing the hairpin-capped ends of the intermediate structures (Lobachev et al., 2002). A recombination assay showed that mutants in *MRE11*, *RAD50*, *XRS2*, and *SAE2* (which controls the endonuclease activity of Mre11 in hairpin processing) exhibited decreased rates of homologous recombination due to their inability to process the hairpin caps of the intermediates (Lobachev et al., 2002). Because the hairpin-capped ends are not processed, homologous ends are not generated, and recombination cannot occur. Fluctuation tests showed that wild-type strains with inverted *Alu* repeats showed an 850-fold increase in recombination rates compared to wild-type strains with direct *Alu* repeats (which cannot form secondary structures). However, in Δ*mre11* and Δ*sae2* mutants, there was only about a 12-fold and 22-fold increase (respectively) in strains with inverted repeats compared to direct repeats.

Inverted repeats can induce gross chromosomal rearrangements, some of which are often seen in tumor cells. Hairpin-capped broken molecules can be converted to inverted dimers as shown in Figure 2 (Lobachev et al., 2002; Narayanan et al., 2006). These inverted dimers can result in gene amplification (extrachromosomal and intrachromosomal) and translocation events. The hairpin processing activity of the MRX complex and Sae2 can prevent the occurrence of such dimers. This shows the importance of the MRX complex and Sae2 in monitoring the genome for hairpin-capped ends, thereby curtailing gross chromosomal rearrangements in the cell.

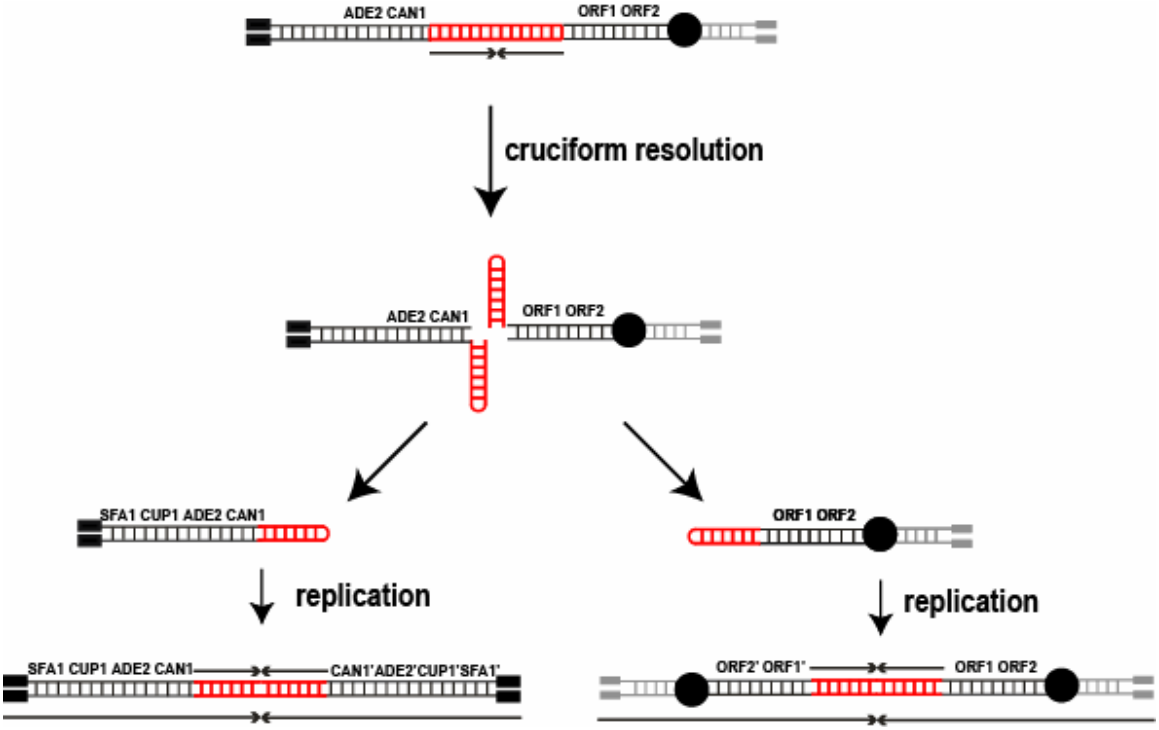

Fig. 2. Model for the formation of dimers. Cruciform resolution leads to the formation of hairpin-capped DNA. Replication of the acentric fragment results in an extrachromosomal gene amplification event, forming a dimer. Replication of the chromosome also leads to the formation of a dimer, resulting in a dicentric chromosome. Figure adapted from (Narayanan et al., 2006).

Consequences of Secondary Structures: Relation to Human Disease

Many different human diseases have been associated with the presence of repetitive sequences, and one prime example is cancer. Palindromes have been found to occur frequently in many different types of cancer cells (Tanaka et al., 2005). These repetitive sequences are not only widespread, but are also distributed in a nonrandom fashion throughout the genomes of different cancer cell lines. While the question can be asked whether these sequences *cause* the disease or are the *consequences* of the disease, there has been some evidence supporting the case of the former in recent years.

Using a yeast system, Narayanan et al., have shown that it is possible to model the events leading to the DNA abnormalities seen in cancer cells (Narayanan et al., 2006). As

described by the authors, it was demonstrated that hairpin-capped intermediates resulting from cruciform resolution can lead to intrachromosomal and extrachromosomal gene amplification events as well as gross chromosomal rearrangements (Fig. 2). The study by Narayanan et al. also identifies rules dictating specific patterns of rearrangements, predicts the location of the causative secondary structure-forming sequences, and might help to uncover susceptible phenotypes. If the mechanism seen in yeast is conserved in all eukaryotes, these rules and predictions can be extended to study cancer genomics.

Clarifying the Mechanism of DSB Formation at the Site of Inverted Repeats

One of the biggest questions concerning the cruciform resolution model described above is the identity of the putative resolvase. In this model, cleavage of the cruciform structure is coupled with the ligation of the nicks to form hairpin-capped molecules. Two distinct enzymes may be involved in this process, or one enzyme could carry out both activities. Several such bifunctional enzymes are known, and can be classified into two classes. The first class includes enzymes related to tyrosine recombinases and type 1B topoisomerases found in prokaryotic organisms and phages that replicate linear genomes with hairpin-capped ends in the place of telomeres (Lobachev et al., 2007, in press and references therein). These genomes carry out a process called telomere resolution using these enzymes. In the second class of enzymes, hairpin-capped molecules are formed as an intermediate during breakage and joining reactions. Some examples of enzymes in this second class include prokaryotic and eukaryotic transposases, retroviral integrases, and the eukaryotic V(D)J recombinase, Rag1. While it is possible that two distinct enzymes

carry out the cruciform resolution process, it cannot be ruled out that the resolvase is a bifunctional enzyme analogous to the enzymes in either of these two classes.

In an attempt to better understand the mechanism of DSB formation at the site of inverted repeats (and possibly lead to the identification of the resolving enzyme), mutant screens for suppressors of the 'Δ*mre11*/Δ*sae2* effect' were performed. As mentioned above, a homologous recombination assay was used to show that mutations in *MRE11* or *SAE2* (in the presence of inverted repeats) lead to a decrease in recombination rates compared to wildtype strains. This result was seen due to the fact that Δ*mre11*/Δ*sae2* mutants cannot process hairpin-capped broken molecules to provide homologous ends for recombination. In these screens, mutants that showed increased recombination rates (comparable to the wild-type level of recombination) in the Δ*mre11* or Δ*sae2* genetic background were isolated. These suppressor mutations could provide further insight into different processes and enzymes involved in the mechanism. For example, it is possible that the increased recombination rates are a result of increased DSBs at the increased number of secondary structure formed. Mutations that allow more opportunities for hairpin or cruciform structures to form would thus provide more substrates for nucleases that can cleave these structures. Alternatively, the increased recombination rates could be due to a mutation in the ligase activity of the resolvase (or the separate ligase involved). In this case, MRX/Sae2 would not be needed to process the hairpin-capped intermediates, and recombination rates would not be affected by Δ*mre11* or Δ*sae2*.

In this study, we used tetrad analysis to map the Δ*mre11* and Δ*sae2* suppressor mutations. Four mutants, all isolated independently, were found to have mutations in different regions of *CDC2* (*POL3*). One of the mutants, *pol3-6PL*, was introduced into

strains containing inverted or direct *Alu* repeats, and an assay was used to look at the rates of arm loss and gene amplification events. Different techniques such as CHEF and Southern blot hybridization were used to further investigate the consequences of the breaks in p*ol3-6PL* cells.

RESULTS

Tetrad Analysis Results Locate Mutation to CDC2 (POL3)

Previously, EMS mutagenesis was performed on Δ*mre11* and Δ*sae2* strains containing inverted repeats to identify suppressor mutations. In addition to the genotypes and phenotypes of the parent strain, many of the resulting mutants were also found to be temperature sensitive (ts) at 37°C or mms sensitive (Lobachev, unpublished data). We characterized one such mutant, RM53, which showed elevated levels of homologous recombination and was also ts.

Tetrad analysis and chromosome walking techniques were used to map the location of the mutation. First, yeast knockout (YKO) strains carrying centromere-linked disruptions, YEL003W::KmX and YHL002W::KmX, were used to determine the distance between the mutated gene and the centromere. The resulting types of tetrads are listed in Table 1. The following equation was used to calculate the distance between the mutated gene and the centromere:

$$\frac{\#\,of\ Tetratypes}{Total\,\#\,of\ Tetrads} \times 50 = \_\_\_\ cM$$

After determining that the gene of interest was located approximately 36 cM away from the centromere, YKO strains containing a disruption near this distance was chosen for each arm of the 16 chromosomes. If the marker and the gene of interest are located

very close to each other on the same arm of the chromosome, these two genes would be linked, single crossing-over events would not occur frequently between them, and double crossovers would be suppressed due to interference. Therefore, most of the resulting tetrads would be parental ditypes (PD), and essentially no non-parental ditypes (NPD) would be seen, since this class comes from double crossing-over events. Tetratypes (T) that result from single crossing-over events can be used to calculate the genetic distance between markers. Results from matings with YDO094C::KmX are shown in the Table 1. The distance between the YDL094C::KmX and the mutated gene was found to be approximately 6 cM.

	YEL003W::KmX	YHL002W::KmX	YDL094C::KmX
NPD	9	10	0
PD	10	10	47
T	53	49	6
Total	72	69	48
Distance (cM)	36.81	35.51	6.25

Table 1. Tetrad analysis data and calculated distances. YKO strains YEL003W::KmX, YHL002W::KmX, and YDL094C::KmX were crossed with the mutant strain, RM53, and the resulting diploid strains were sporulated. YEL003W::KmX and YHL002W::KmX were used as centromere-linked markers. Tetrad dissection resulted in the numbers of non-parental ditypes (NPD), parental ditypes (PD), and tetratypes (T) listed above. Using this data, the distance between the marker gene (or centromere) and the mutant gene was calculated for each.

Finally, chromosome walking techniques were used to identify the mutated gene. YKO strains with disrupted ORFs near 36 cM from the centromere on the left arm were PCR amplified to create overlapping fragments along this region of the chromosome (Fig. 3). These PCR products containing the KmX cassette were then transformed into the mutant strain. Media containing G418 was used to select for integration events. If the wt copy of the mutated gene was present on the PCR product, the mutant sequence can be

converted into the wt sequence during integration as a result of recombination events between the chromosome and homologous regions in the fragment. A high frequency of the disappearance of the ts phenotype among the KmX transformants would indicate the presence of the complementing sequence in the PCR product. Using this method, *CDC2* (*POL3*) was found to be the mutant gene.

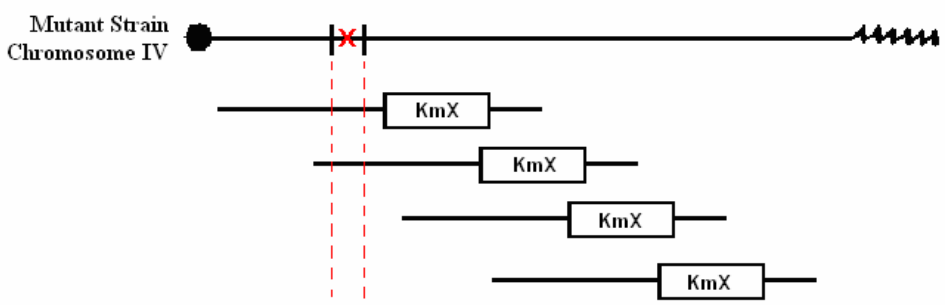

Fig. 3. Chromosome walking techniques. To identify the location of the mutant gene, YKO strains containing kanamycin disruptions near 36 cM from the centromere on the left arm were PCR amplified and transformed into mutant strains. Due to homology, these PCR products are then integrated into the chromosome, replacing the existing region of DNA. The kanamycin disruption was used to select for these integration events on G418 media. If the wt copy of the mutant gene is present on the PCR product, the mutation will be replaced during integration, and strains will no longer be ts.

In addition to RM53, three other mutants found were allelic to RM53. DNA sequencing results showed that the mutations were not localized to a specific domain within the *POL3* gene, but rather scattered throughout (Fig. 4). The amino acid change resulting from these mutations are shown in Table 2.

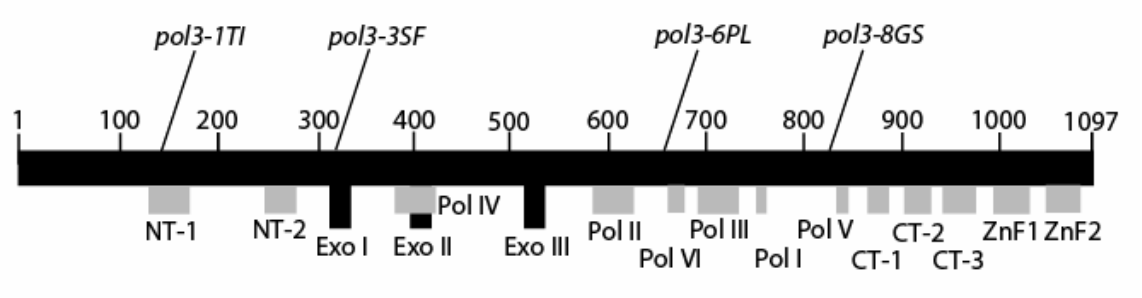

Fig. 4. Locations of the mutations found in *POL3*. The 1097 aa, *POL3* ORF is shown above. The locations of the major domains and conserved regions are identified with gray and black shaded boxes. The locations of the mutations found are depicted as well. Figure adapted from (Jin et al., 2001).

Strain	Allele	Change
RM53	*pol3-3SF*	S319F
HM1	*pol3-6PL*	P664L
RM51a	*pol3-1TI*	T143I
RM62	*pol3-8GS*	G834S

Table 2. Mutant strains and their amino acid changes. The four mutant alleles of *POL3* that were found in this study are listed in this table. The amino acid change resulting from point mutations in the gene (as well as their positions) have been identified.

Experimental System: Arm Loss / Gene Amplification Assay

To further investigate the consequences of these mutations in the presence of inverted repeats, the following construct was used. This experimental system utilizes the fact that chromosome V contains a 43 kb region on its left arm that only contains nonessential genes (Narayanan et al., 2006). This region can therefore be lost, and the cells would still be viable. A *LYS2* gene has been placed in this region, and disrupted with the insertion of inverted or direct *Alu* repeats (Fig. 5). Inverted repeat sizes of 320 bp, 180 bp, 120 bp, 70 bp, and 40 bp were used, and all contained a 12 bp spacer. 320 bp direct *Alu* repeats were used as a control, due to the fact that these sequences cannot form secondary structures.

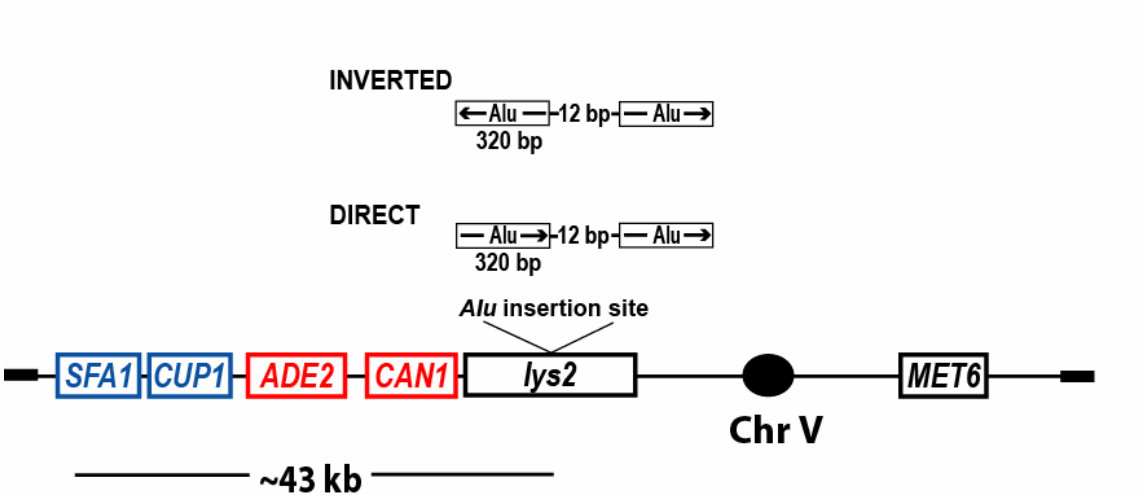

Fig. 5. Experimental System used to Study Inverted-Repeat Induced Chromosomal Instability. Chromosome V, which possesses a 43 kb region on its left arm that contains only nonessential genes, was used to track arm loss events (*ADE2* and *CAN1* markers) and gene amplification events (*SFA1* and *CUP1* markers). Inverted *Alu* repeats of different sizes containing a 12 bp spacer, were inserted into the *LYS2* gene in this region. Direct *Alu* repeats (320 bp with a 12 bp spacer) was used as a control. Figure adapted from (Narayanan et al., 2006).

To follow arm loss events, the counterselectable marker, *CAN1*, was used. *CAN1* codes for an arginine permease that allows cells to uptake canavanine, a toxic analog of arginine. Cells would be able to survive only if *CAN1* was inactivated by a mutation, or if the gene was lost. To select for the second event, *ADE2* was used as a dual marker. This gene product is involved in adenine biosynthesis, and cells lacking this gene would appear red in color. Cells that have undergone an armloss event would be Can^R Ade^-.

To follow gene amplification events, two gene dosage markers were used: *SFA1* and *CUP1*. *SFA1* codes for a formaldehyde dehydrogenase, and *CUP1* codes for a copper chelating enzyme. Therefore, when selection is applied, cells that have amplified these genes would be able to survive in higher concentrations of formaldehyde and copper, which are normally toxic for the cell at higher concentrations.

For the studies described in the remainder of this paper, we have focused on one specific mutant allele, *pol3-6PL*.

Increased Rates of Arm Loss and Gene Amplification in pol3-6PL Mutants

Site-directed mutagenesis was used to introduce the *pol3-6PL* mutation into strains with the above construct. Fluctuation tests were performed to look at the rates of arm loss and gene amplification events. The resulting data is shown in Table 3.

Insertion in *LYS2*	Repeat Size (bp)	Arm Loss Events ($x10^7$)		Gene Amp. Events ($x10^7$)	
		wt	*Pol3-6PL*	wt	*Pol3-6PL*
Direct *Alus*	320	0.015 (0.010-0.020)	0.444 (0.175-1.535)	0.02 (0.01-0.03)	0.128 (0.068-0.220)
Inverted *Alus*	320	371 (294-421)	11948 (9801-13208)	229 (185-316)	588 (422-735)
	180	59.47 (55.76-73.50)	1410 (1211-1594)	17.50 (15.59-24.84)	53.69 (43.95-57.68)
	120	6.99 (6.37-10.30)	159 (136-214)	5.88 (3.96-8.15)	9.53 (9.18-11.82)
	70	8.70 (6.31-12.14)	40.89 (34.13-47.79)	8.67 (8.20-9.04)	12.43 (11.01-13.83)
	40	9.13 (7.57-10.48)	4.17 (2.71-5.34)	4.20 (3.72-5.64)	5.43 (2.41-15.98)

Table 3. Rates of arm loss and gene amplification events. Rates of arm loss and gene amplification events in the presence of direct and inverted *Alu* repeats for wt strains as well as *pol3-6PL* strains. For the repeats in the direct orientation, only the 320 bp sized repeat was used. For the repeats in the inverted orientation, repeats of different sizes were used. The numbers in parentheses show the 95% confidence interval for each of the rates. This fluctuation data was obtained using the chromosome V experimental system described in the previous section.

As expected, the rates for both the arm loss events and gene amplification events were higher for strains containing inverted *Alu* repeats than directed *Alu* repeats (control). In wt strains, there was a 25,000-fold increase in arm loss events for 320 bp inverted *Alus*

vs. 320 bp direct *Alus*. For gene amplification events in wt strains, there was about an 11,450-fold increase. In addition, as the size of the inverted repeats decreases, the rates generally decrease as well. Similar results were observed for *pol3-6PL* strains.

Next, comparisons of wt strains to *pol3-6PL* strains show large fold increases as well (Table 4). In strains containing 320 bp direct *Alu* repeats, there was a 30-fold increase in the rate of arm loss events for *pol3-6PL* compared to wt. Similarly, for 320 bp inverted *Alu* repeats, there was a 32-fold increase. For gene amplification events, there was only about a 6.4-fold increase and 2.6-fold increase in *pol3-6PL* vs. wt strains for direct and inverted *Alus* respectively.

Insertion in *LYS2*	Repeat Size (bp)	Fold Difference (*pol3-6PL/ wt*)	
		Arm Loss Events	Gene Amp. Events
Direct *Alus*	320	30	6.4
Inverted *Alus*	320	32	2.6
	180	24	3
	120	23	1.6
	70	4.7	1.4
	40	0.5	1.3

Table 4. Fold difference for *pol3-6PL* over wt strains. Comparison of *pol3-6PL* strains vs. wt strains for each of the repeat sequences show fold differences listed above. Arm loss events in direct *Alu* repeats show a 30-fold increase. The rates for inverted *Alus* show similar fold increases for repeat sizes of 320 bp, 180 bp, and 120 bp. Inverted *Alus* containing repeats of 70 bp and 40 bp show much smaller fold differences. Overall, gene amplification events showed much smaller fold increases across the board.

Visualizing the DSB using Southern Blot Hybridization

In order to determine whether the increased armloss and gene amplification events were the result of increased breakage at the inverted repeats, the amount of DSBs

in *pol3-6PL* was detected using pulse-field gel electrophoresis (specifically CHEF) and Southern blot hybridization. After separating the chromosomes using CHEF, Southern blot hybridization was performed using a ^{32}P labeled probe to visualize the DSB formation (Fig. 6.). The probe was homologous to a region within *HPA3*, located on the left arm of chromosome V. Lane 2 contains wt strains with 320 bp direct *Alus*. For these strains, only one band at 586 kb corresponding to the intact chromosome V is present. Lanes 3-5 contain wt, Δ*sae2*, and Δ*mre11* strains respectively (all with 320 bp inverted *Alus*). In addition to the 586 kb band corresponding to chromosome V, bands are also present at 43 kb (break) and 86 kb (inverted dimer), but are very weak. Lanes 6-8 contain *pol3-6PL*, *pol3-6PL*+Δ*sae2*, and *pol3-6PL*+Δ*mre11* strains respectively (all in the presence of 320 bp inverted *Alus*). Once again, chromosome V is strongly visible. Bands corresponding to the break and inverted dimer are plainly visible as well. For *pol3-6PL* (lane 6), the band corresponding to the broken fragment at 43 kb is shifted upwards, indicating that these fragments are running more slowly. The several fragments seen in all lanes between the 97 kb band and 586 kb band are due to unspecific hybridization of the probe.

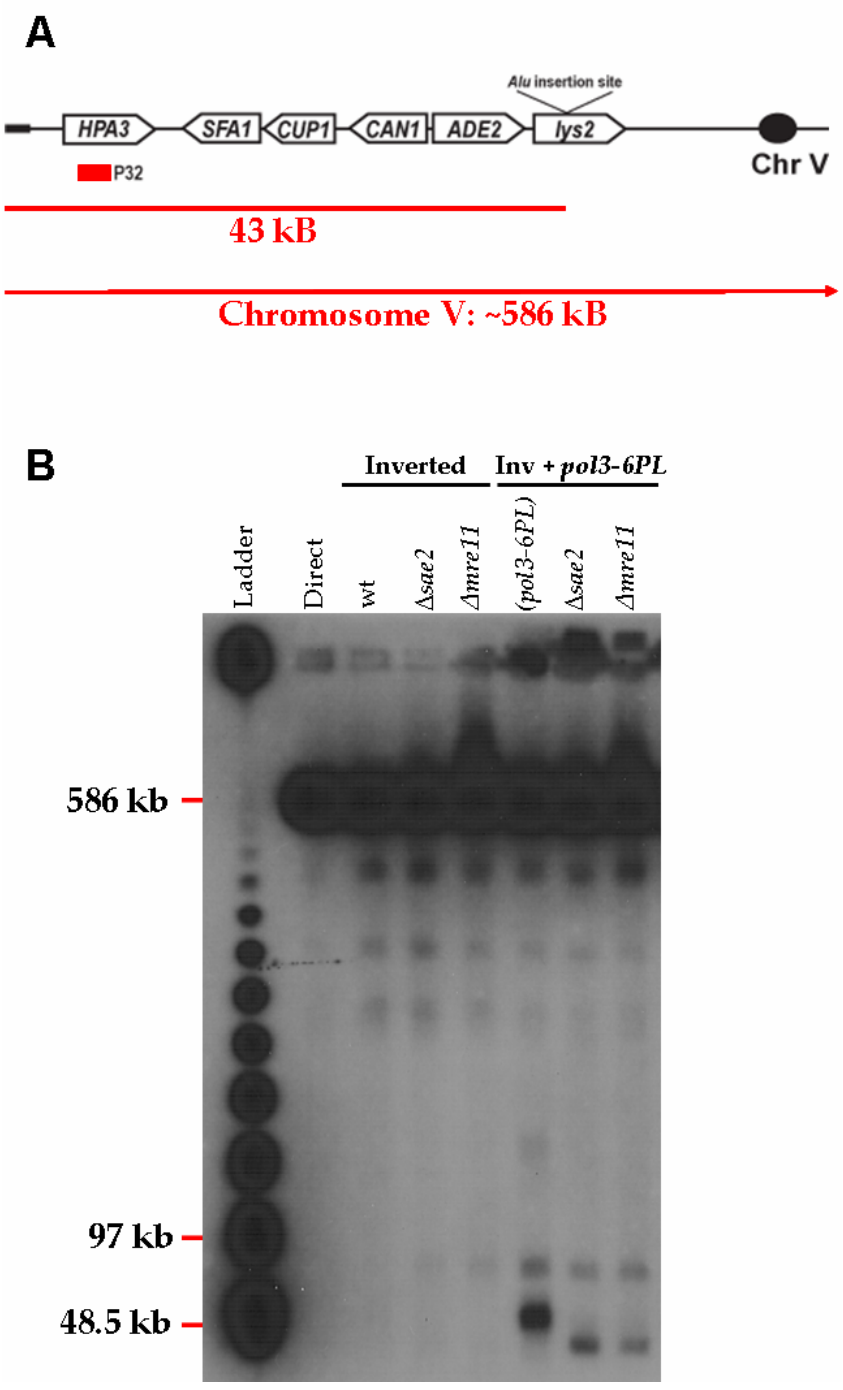

Fig. 6. Detection of DSBs using CHEF and Southern Blot Hybridization. (A) The ^{32}P labeled probe used for Southern Blot Hybridization is complementary to a region within *HPA3*, located on the left arm of chromosome V. Intact chromosome V would run as a 586 kb fragment. In the event of a DSB, the 43kb broken fragment would be detected since the probe hybridizes to the centromere distal side of the break. (B) Chromosomes were separated using a CHEF gel, and a ^{32}P labeled probe was used to detect chromosome V. Lane 1 contains the λ ladder probed with λ-specific probes. Lane 2 contains wt strains with 320 bp direct *Alus*. Lanes 3-5 contain *POL3*$^{+}$ strains with 320 bp inverted *Alus* (with or without additional mutations as noted). Lanes 6-8 contain *pol3-6PL* strains with 320 bp inverted *Alus* (with or without additional mutations as noted). The 43 kb bands correspond to the broken fragment resulting from DSB formation on chromosome V, while the 86 kb bands correspond to the dimerization of this fragment.

DISCUSSION

Polymerase δ (POL3)

The presence of 320 bp inverted repeats in wt strains lead to an 850-fold increase in the rate of recombination compared with wt strains with 320 bp direct repeats (Lobachev et al., 2002). If a mutation was introduced into *MRE11* or *SAE2*, the rate of recombination was only about 12-fold and 22-fold higher respectively. This decrease in recombination rates is thought to be due to the inability of the MRX complex to process hairpin-capped intermediates that formed as a result of DSBs. Since these hairpins are not being processed, homologous ends needed for recombination are not being generated. Mutants that suppressed this Δ*mre11*/Δ*sae2* effect were found using EMS mutagenesis. Recombination rates in these mutant strains were found to be comparable to wt levels (Lobachev et al, unpublished data).

Using tetrad analysis and chromosome walking techniques, the mutation in the RM53 strain was mapped to *POL3*. The *POL3* gene codes for the 125 kDa essential catalytic subunit of polymerase δ, one of the three main replicative polymerases in eukaryotes (Johansson et al., 2001). Pol3, along with Pol31 (an essential 48 kDa subunit) and Pol32 (a non-essential 55 kDa subunit), make up the monomeric unit of Pol δ. Pol δ is highly conserved in eukaryotes and contains a 3' to 5' exonuclease proofreading activity. In addition, it has been shown that the activity of RFC (a clamp loader) and PCNA (which functions as a sliding clamp) is needed to tether Pol δ to the DNA (Bambara et al., 1997).

Three other mutants, all isolated independently, were found to be allelic to RM53 as well. The four mutants found in *POL3* are not localized in a specific domain, but are

distributed throughout the gene (Fig. 4). This suggests that these mutations are affecting the function of Pol δ by changing the tertiary/quaternary structure of the protein. This decreased function of Polδ is somehow leading to an increase in the recombination rates observed in the Δ*mre11* and Δ*sae2* background. Also, the fact that the mutation is found in a polymerase strongly suggests a replication-dependent mechanism of breakage.

Increased Frequency of Breaks in pol3-6PL Acts to Suppress the Δmre11/Δsae2 effect

Using the arm loss / gene amplification assay system described above, fluctuation tests were performed on wt and *pol3-6PL* strains containing inverted and direct *Alu* repeats (Table 3). 320 bp direct *Alu* repeats were used as the control due to the fact that direct repeats cannot form secondary structures. When inverted *Alu* repeats were used, there was an increase in arm loss and gene amplification events in all strains compared to the control. This supports the idea that the presence of inverted repeats leads to an increased frequency of breakage, presumably from the formation of secondary structures. The data also shows that the frequency of these events decreases as the size of the repeat decreases. This agrees with previous findings that decreasing the size of repeats decreases stability of the repeats in the formation secondary structures (Lobachev et al., 1998).

The fluctuation test data can be used to look at the effect of the *pol3-6PL* mutation in comparison to wt strains. The fold difference in arm loss events between *pol3-6PL* strains and wt strains are shown in Table 4. For 320 bp inverted *Alus*, there was a 32-fold increase in *pol3-6PL* versus wt strains. Similar fold increases were shown for 180 bp and 120 bp inverted *Alus*. It is interesting to note that the effect of *pol3-6PL* starts to decrease drastically once the size of the inverted repeat reaches 70 bp and 40 bp. This

may be due to the fact that these smaller repeats form less stable secondary structures, and therefore the probability of DSB formation (and arm loss events) are less likely to occur. Looking at both Table 3 and Table 4 however, it is evident that there is a drastic increase in arm loss events (and therefore DSBs) in the presence of *pol3-6PL*.

One hypothesis for this observed increase in breaks is that the unwinding of the DNA by the helicase at the replication fork may be uncoupled with DNA synthesis, allowing for a larger region of unwound DNA. This would allow for cruciform structures to form more easily, and therefore more substrates would be available for enzymes involved in resolving these structures. In the absence of functional MRX complexes, the hairpin intermediates that are formed after cruciform resolution may be processed by alternative enzymes, resulting in the increased recombination rates observed in the preliminary studies mentioned above.

Southern Blot Hybridization Shows Break and Dimer in POL3 mutants

Southern Blot Hybridization revealed that strains with direct *Alu* repeats had only one band, which corresponds to the intact chromosome V (Fig. 6). This is expected since direct repeats are unable to form secondary structures, and therefore would not lead to DSB formation at this location. In the presence of 320 bp inverted *Alu* repeats, strains lacking *pol3-6PL* showed weak bands corresponding to the break and inverted dimer. These bands are weak due to the fact that DSBs occur at much lower rates without this additional mutation in *POL3* (as seen from fluctuation test data).

Strains with *pol3-6PL* in the presence of inverted repeats clearly show the break and inverted dimer. For *pol3-6PL*+Δ*sae2* and *pol3-6PL*+Δ*mre11*, the break and dimer

appear at the expected lengths of 43 kb and 86 kb respectively. However, while the inverted dimer is running at the expected length in *pol3-6PL*, the band corresponding to the break is running slower in this strain. It is thought that this is due to processing of the hairpin-capped end of the broken fragment by the functional MRX complex. Once these fragments are processed, single-stranded ends of variable lengths are generated, causing hindrance as the fragment runs through the gel. Therefore, the band corresponding to these fragments would be observed to be running more slowly in the gel.

In addition, the bands corresponding to the broken fragment and inverted dimer for *pol3-6PL* are much brighter than the same bands for *pol3-6PL*+Δ*sae2* and *pol3-6PL*+Δ*mre11*. This observation is consistent with additional fluctuation tests performed for *pol3-6PL*+Δ*sae2* strains in the presence of 320 bp inverted *Alus* (data not shown). These tests demonstrated a 23-fold increase in arm loss events over wt strains. While this is still a large fold increase, it is lower than what is observed for *pol3-6PL* alone (32-fold increase). Increase in arm loss events would correspond to the increase in the 43 kb broken fragment seen on the Southern gel. This observation is explained in further detail in the following sections.

2D Gel Electrophoresis Revealed Two Sites of Replication Arrest

Due to the fact that the suppressor mutation was found in *POL3*, it was strongly suspected that compromised replication was involved in the mechanism for DSB formation at the site of inverted repeats. Therefore, Lobachev et al. performed 2-dimensional gel electrophoresis to look for replication arrest sites (unpublished data). The replication of a 4.5 kb fragment containing the 320 bp inverted *Alu* repeats was further

investigated. 2D gel electrophoresis was used to first separate the DNA in the first dimension by mass, then in the second dimension by shape. If there is replication arrest, a stronger signal will be seen at the position of the arc corresponding to the structure of the stalled replication fork. In wild type strains, replication arrest is barely visible on the arc. However, in *pol3-6PL*, two distinct regions of replication arrest can be seen on the arc.

New Model of Replication-Dependent Mechanism of Chromosomal Instability

The 2D gel electrophoresis data (Lobachev et al, unpublished data) shows the presence of two replication arrest intermediates. From this data and above information, we propose that there are two different pathways that can lead to the genomic instability described above (Fig.7). During replication, cruciform formation can occur in the DNA due to the relaxed supercoiling of the duplex DNA. In the cruciform resolution pathway (pathway I), the replication fork reaches the cruciform structure and stalls. The cruciform is then cleaved and hairpin-capped ends are generated. This leads to the armloss and gene amplification events that have been observed. In *pol3-6PL* mutants, cruciform structures may occur more frequently (due to uncoupled helicase-polymerase activities), leading to the observed increase in arm loss and gene amplification events.

In the hairpin formation pathway (pathway II), the template DNA in the lagging strand remains single-stranded longer than the leading strand, allowing the inverted repeat to base-pair within itself to form a hairpin structure. This blocks the replication fork and causes it to stall. The MRX complex (and possibly other enzymes) is involved in processing the hairpin structure. Again, *pol3-6PL* mutants may synthesize the lagging strand DNA more slowly, allowing for increased hairpin formation.

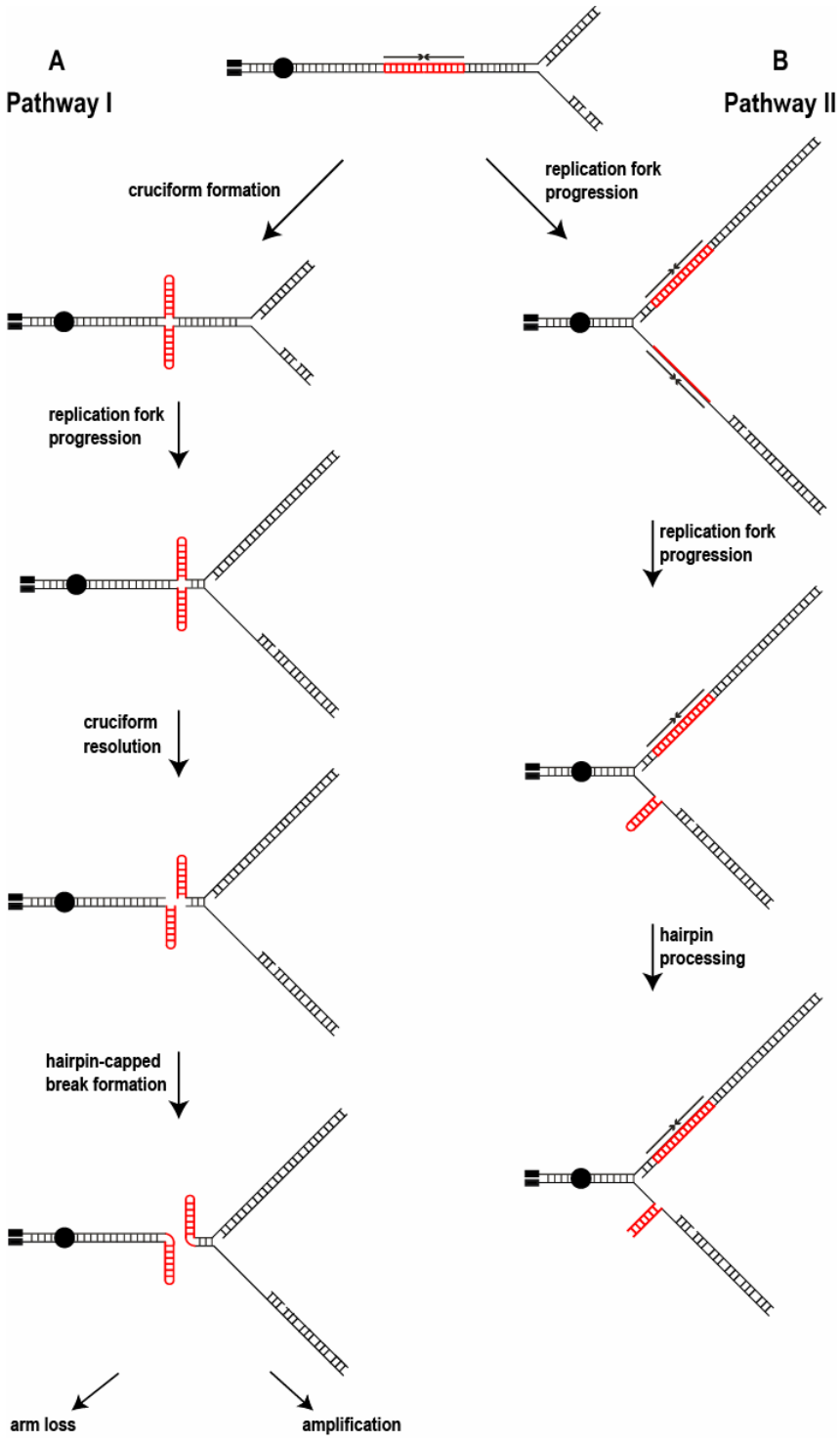

Fig. 7. Proposed model for replication-dependent chromosome instability. (A) In pathway I, inverted repeats form cruciform structures during replication due to unwinding of the supercoiled DNA ahead of the replication fork. As replication progresses, the replication fork reaches the cruciform and stalls. Cruciform resolution leads to arm loss and gene amplification events. (B) In pathway II, the replication fork progresses through the inverted repeats and cruciform structures do not form. However, due to a larger single-stranded region in the lagging strand, the exposed inverted repeat forms a hairpin structure, which can be processed by the MRX complex.

CHEF and Southern Blot Hybridization data from Figure 6 shows an increased intensity of the bands corresponding to the broken fragments and inverted dimers in *pol3-6PL* (compared to *pol3-6PL*+Δ*sae2* and *pol3-6PL*+Δ*mre11*). We believe that this is due to the fact that in the presence of the Δ*sae2* (or Δ*mre11*) mutation, the second pathway in the above model is not relevant in the contribution of the formation of broken fragments. Therefore the only broken fragments that result come from the first pathway of the model, through cruciform resolution. For *pol3-6PL*, however, both pathways contribute to the formation of the broken fragment. Cruciform resolution still yields these breaks, and in addition, hairpin processing by the MRX complex in the second pathway would also contribute to the formation of breaks.

Conclusion

In this study, several suppressors for the Δ*mre11/*Δ*sae2* effect were mapped to the *POL3* gene. Here, we focused on *pol3-6PL*. Using the experimental system described above, we found rates of arm loss events and gene amplification events to be higher in *pol3-6PL* strains compared to wild type strains. CHEF and Southern blot hybridization revealed the presence of a 43 kb fragment corresponding to the broken arm, as well as an 86 kb fragment corresponding to a dimer. 2D gel electrophoresis showed not one, but two stalled replication forks in *pol3-6PL* strains. We have therefore proposed a new model in which DSBs form via two pathways. In the first pathway, DSBs are proposed to form via cruciform formation and resolution ahead of the replication fork. In the second pathway, DSBs are proposed to form in the lagging strand via hairpin processing.

Currently, we are looking at *pol3-1TI,* one of the other mutants found to be allelic to *pol3-6PL*. Interestingly, while this mutant shows increased recombination levels, it is not temperature sensitive and does not show any other associated phenotypes. We are interested in further studying this mutant because it may more closely resemble mutations that have a subtle effect which allows the cell to continue functioning almost normally, but in the presence of other mutations or factors, may become more detrimental. In pre-cancerous cells, it is thought that the buildup of subtle effects contributes to the cell's transformation into a cancerous cell. Further investigation may shed more light onto the processes of disease development.

In addition, this experimental system provides an excellent way to look at the effects of human small nucleotide polymorphisms (SNPs) found in human cancer cells. Because the gene for the catalytic subunit of pol δ is highly conserved, it is possible to align and compare the yeast *POL3* gene and the human *POLD1* gene. Amino acid changes corresponding to the SNPs found in *POLD1* can be introduced into our yeast system to monitor the effects. This would provide a way to investigate and compare the mechanism of genomic instability in humans.

ACKNOWLEDGEMENTS

We thank H. Kim, M. Medrzycki, J. Tucker, S. Black, T. Bodrogi, G. Lasker, and H. Gong for their assistance. We would especially like to thank V. Narayanan for her guidance and help throughout this work. This research was supported by funding from PURA (2005) and URS (2006).

MATERIALS AND METHODS

Strains

All strains used in this study were isogenic to KS520 (*MAT*a*, his7-2, leu2-3,112, trp1*-Δ*, ura3*-Δ*, lys2*-Δ*, ade2*-Δ*, bar1*-Δ*, sfa1*-Δ*, cup1-1*-Δ*, yhr054c*-Δ*, cup1-2*-Δ). *SFA1*, *CUP1*, *ADE2*, and *LYS2* genes were inserted into the left arm of chromosome V, between 34,211 bp and 34212 bp (Narayanan et al., 2006, supplemental information).

Yeast knockout (YKO) strains from Open Biosystems were used in matings for tetrad analysis. The specific library used for our studies was created from yeast parental strain BY4742 (*MAT*α*, his3*Δ*1, leu2Δ0, lys2Δ0, ura3*Δ*0*). Each of the strains in the library contained a nonessential gene disrupted with the *kanMX cassette.*

Tetrad Analysis & Chromosome Mapping

RM53 was mated with strains from the YKO library described above. The resulting diploid strain was sporulated on sporulation media (0.1% yeast extract, 0.98% potassium acetate, 0.05% dextrose), and tetrad dissection was then performed using Singer microscope systems.

Centromere-linked YKO strains (YEL003W::KmX & YHL002W::KmX) were used first to determine the distance of the mutant gene from the centromere. After this distance was determined, YKO strains containing a disruption near this distance was chosen for each arm of the 16 chromosomes. These 32 strains were to be used to mate with RM53 to look for linkage. YDL094C::KmX was used to determine that the gene of interest was located on chromosome IV.

After the chromosome arm (and region) containing the mutant gene was determined, chromosome walking techniques were used to determine the gene of interest. Different regions of chromosome IV near YDL094C were PCR-amplified from the YKO library. These PCR products were then transformed back into RM53, and complementation of the ts phenotype was selected for.

Site-directed mutagenesis

Integrative plasmids containing either the N-terminal (YIpKhr5) or C-terminal (p170) region of *POL3* were obtained from D. Gordenin (NIH/NIEHS, Research Triangle Park, NC). The plasmids were used as a template for site-directed mutagenesis using the Quikchange Site-Directed Mutagenesis Kit (Stratagene), as directed by the manufacturer. Primers containing the desired mutations were used in PCR amplification with *Pfu* polymerase, thereby incorporating the desired mutations into the gene of interest. DpnI was used to digest the non-mutated parental DNA before bacterial transformation into XL1-Blue supercompetant cells. After verification of the plasmid sequence, correct plasmids were transformed into desired yeast strains (described above). Nucleotide sequences of these primers are available on request.

Genetic Techniques

Genetics techniques were performed as previously described in (Lobachev et al., 2000). Rates of arm loss and gene amplification and 95% confidence levels were determined by fluctuation tests using at least 14 independent cultures.

CHEF and Southern Blot Hybridization

Pulse field gel electrophoresis (using contour-clamped homogeneous electric field, or CHEF) was performed to separate chromosomes. Chromosomal DNA was embedded into agarose plugs using the CHEF Genomic DNA Plug kit from Bio-Rad. Gels were run in 0.5X TBE at 14°C using the Bio-Rad CHEF Mapper XA for 30 hours with switch times of 2.91 sec – 48.53 sec.

A ^{32}P labeled probe that is complementary to a region within *HPA3* (located on the left arm of chromosome V), was used for southern blot hybridization. Southern blot hybridization was performed as previously described in (Lobachev et al., 2002). The nucleotide sequence of the primers used to generate the fragments for labeling are available on request.

References

Bambara, R.A., R.S. Murante, and L.A. Henricksen. 1997. Enzymes and Reactions at the Eukaryotic DNA Replication Fork. *J. Biol. Chem.* 272:4647-4650.

Bissler, J.J. 1998. DNA Inverted Repeats and Human Disease. *Frontiers in Bioscience*. 3:408-418.

Connelly, J.C., E.S. de Leau, and D.R.F. Leach. 1999. DNA cleavage and degradation by the SbcCD protein complex from *Eschericia coli*. *Nucleic Acids Res.* 27:1039-1046.

Connelly, J.C., and D.R.F. Leach. 1996. The *sbcC* and *sbcD* genes of *Escherichia coli* encode a nuclease involved in palindrome inviability and genetic recombination. *Genes to Cells*. 1:285-291.

Jin, Y.H., R. Obert, P.M.J. Burgers, T.A. Kunkel, M.A. Resnick, and D. Gordenin. 2001. The 3' to 5' exonuclease of DNA polymerase delta can substitute for teh 5' flap endonuclease Rad27/Fen1 in processing Okazaki fragments and preventing genome instability. *PNAS*. 98:5122-5127.

Johansson, E., J. Majka, and P.M.J. Burgers. 2001. Structure of the DNA Polymerase delta from *Saccharomyces cerevisiae*. *J. Biol. Chem.* 276:43824-43828.

Leach, D.R.F. 1994. Long DNA Palindromes, Cruciform Structures, Genetic Instability and Secondary Structure Repair. *Bioessays*. 16:893-900.

Leach, D.R.F., and F.W. Stahl. 1983. Viability of Lambda-Phages Carrying a Perfect Palindrome in the Absence of Recombination Nucleases. *Nature*. 305:448-451.

Lewis, S., E. Akgun, and M. Jasin. 1999. Palindromic DNA and genome stability - Further studies. *In* Molecular Strategies in Biological Evolution. Vol. 870. 45-57.

Lewis, S., and A.G. Cote. 2006. Palindromes and genomic stress fractures: Bracing and repairing the damage. *DNA Repair*. 659:15.

Lloyd, R.G., and C. Buckman. 1985. Identification and Genetic Analysis of *sbcC* Mutations in Commonly Used *recBC sbcB* Strains of *Escherichia coli* K-12. *J. Bacteriol.* 164:836-844.

Lobachev, K.S., D.A. Gordenin, and M.A. Resnick. 2002. The Mre11 complex is required for repair of hairpin-capped double-strand breaks and prevention of chromosome rearrangements. *Cell*. 108:183-193.

Lobachev, K.S., A. Rattray, and V. Narayanan. 2007. Hairpin- and Cruciform-Mediated Chromosome Breakage: Causes and Consequences in Eukaryotic Cells. *Frontiers in Bioscience*.

Lobachev, K.S., B.M. Shor, H.T. Tran, W. Taylor, J.D. Keen, M.A. Resnick, and D.A. Gordenin. 1998. Factors affecting inverted repeat stimulation of recombination and deletion in Saccharomyces cerevisiae. *Genetics*. 148:1507-1524.

Lobachev, K.S., J.E. Stenger, O.G. Kozyreva, J. Jurka, D.A. Gordenin, and M.A. Resnick. 2000. Inverted Alu repeats unstable in yeast are excluded from the human genome. *EMBO J.* . 19:3822-3830.

Narayanan, V., P.A. Mieczkowski, H.M. Kim, T.D. Petes, and K.S. Lobachev. 2006. The pattern of gene amplification is determined by the chromosomal location of hairpin-capped breaks. *Cell*. 125:1283-96.

Sherman, F. 2002. Getting started with yeast. *Meth. Enzymol.* 350:3-41.

Sinden, R.R. 1994. DNA Structure and Function. Academic Press, San Diego.

Tanaka, H., D.A. Bergstrom, M.C. Yao, and S.J. Tapscott. 2005. Widespread and nonrandom distribution of DNA palindromes in cancer cells provides a structural platform for subsequent gene amplification. *Nat. Genet.* 37:320-327.

Trujillo, K.M., and P. Sung. 2001. DNA Structure-Specific Nuclease Activities in the *Saccharomyces cerevisiae* RAD50-MRX Complex. *J Biol Chem*. 276:35458-35464.